Telecom Churn Case Study

Pragya Bhargava
Anand Prakash

Following are the Top 10 Rows of telecom Churn data set. Showing Mobile number, Circle id....

```
[4]  1 tcc_data.head(10)
```

	mobile_number	circle_id	loc_og_t2o_mou	std_og_t2o_mou	loc_ic_t2o_mou	last_date_of_month_6	last_date_of_month_7	last_date_of_month_8
0	7000842753	109	0.0	0.0	0.0	6/30/2014	7/31/2014	8/31/2014
1	7001865778	109	0.0	0.0	0.0	6/30/2014	7/31/2014	8/31/2014
2	7001625959	109	0.0	0.0	0.0	6/30/2014	7/31/2014	8/31/2014
3	7001204172	109	0.0	0.0	0.0	6/30/2014	7/31/2014	8/31/2014
4	7000142493	109	0.0	0.0	0.0	6/30/2014	7/31/2014	8/31/2014
5	7000286308	109	0.0	0.0	0.0	6/30/2014	7/31/2014	8/31/2014
6	7001051193	109	0.0	0.0	0.0	6/30/2014	7/31/2014	8/31/2014
7	7000701601	109	0.0	0.0	0.0	6/30/2014	7/31/2014	8/31/2014
8	7001524846	109	0.0	0.0	0.0	6/30/2014	7/31/2014	8/31/2014
9	7001864400	109	0.0	0.0	0.0	6/30/2014	7/31/2014	8/31/2014

Detail attributes of Dataset

```
Index(['last_date_of_month_6', 'last_date_of_month_7', 'last_date_of_month_8',
       'last_date_of_month_9', 'date_of_last_rech_6', 'date_of_last_rech_7',
       'date_of_last_rech_8', 'date_of_last_rech_9',
       'date_of_last_rech_data_6', 'date_of_last_rech_data_7',
       'date_of_last_rech_data_8', 'date_of_last_rech_data_9'],
      dtype='object')
(3972, 226)
```

```
1 tcc_data.describe()
```

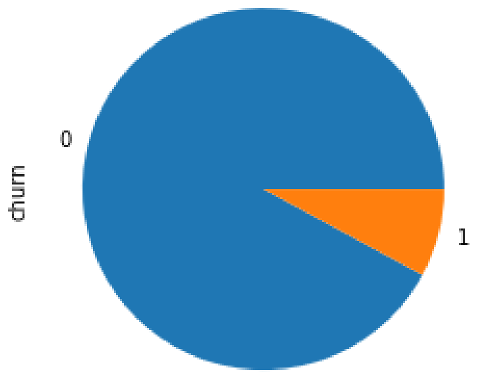
	mobile_number	circle_id	loc_og_t2o_mou	std_og_t2o_mou	loc_ic_t2o_mou	arpu_6	arpu_7	arpu_8	arpu_9	onnet_mou_6	onnet_mou_7	onnet_mou_8	onnet_mou_9
count	3.972000e+03	3972.0	3939.0	3939.0	3939.0	3972.000000	3972.000000	3972.000000	3972.000000	3842.000000	3823.000000	3755.000000	3659.000000
mean	7.001218e+09	109.0	0.0	0.0	0.0	285.045121	278.942359	274.410348	257.626496	131.517657	129.731038	132.414828	132.521110
std	6.924414e+05	0.0	0.0	0.0	0.0	302.244325	318.959997	304.036897	293.650251	299.791531	299.176457	349.503417	350.595283
min	7.000000e+09	109.0	0.0	0.0	0.0	-2041.228000	-2014.045000	-945.808000	-267.243000	0.000000	0.000000	0.000000	0.000000
25%	7.000632e+09	109.0	0.0	0.0	0.0	95.540750	88.581750	80.990000	61.738000	7.760000	6.610000	5.985000	4.770000
50%	7.001223e+09	109.0	0.0	0.0	0.0	203.154500	193.914000	189.894500	176.675500	35.000000	31.410000	31.330000	28.930000
75%	7.001811e+09	109.0	0.0	0.0	0.0	375.462750	367.937000	370.228000	352.587750	114.475000	108.615000	109.535000	109.600000
max	7.002410e+09	109.0	0.0	0.0	0.0	3959.954000	6453.689000	3327.711000	3835.053000	6459.340000	6372.530000	10752.560000	10427.460000

Missing Values in the Datasets

```
#identifying missing values in the dataset
((tcc_data.isnull().sum()/tcc_data.shape[0])*100).round(2).sort_values(ascending=False)
```

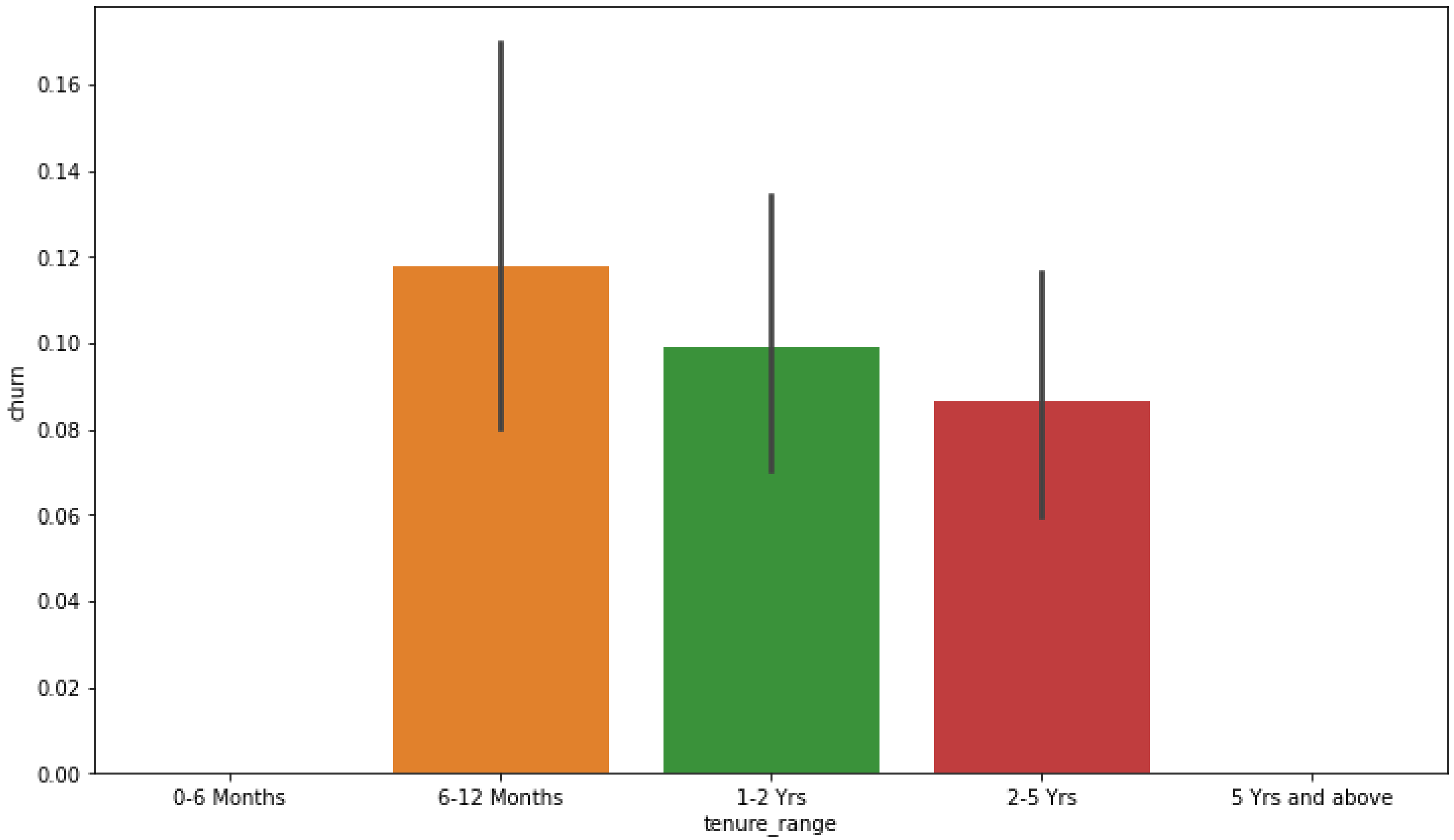
```
fb_user_6                   74.60
count_rech_2g_6             74.60
arpu_2g_6                   74.60
arpu_3g_6                   74.60
max_rech_data_6             74.60
av_rech_amt_data_6          74.60
night_pck_user_6            74.60
date_of_last_rech_data_6    74.60
count_rech_3g_6             74.60
total_rech_data_6           74.60
av_rech_amt_data_9          74.07
date_of_last_rech_data_9    74.07
max_rech_data_9             74.07
count_rech_3g_9             74.07
total_rech_data_9           74.07
count_rech_2g_9             74.07
arpu_2g_9                   74.07
night_pck_user_9            74.07
fb_user_9                   74.07
arpu_3g_9                   74.07
date_of_last_rech_data_7    73.36
night_pck_user_7            73.36
max_rech_data_7             73.36
fb_user_7                   73.36
count_rech_2g_7             73.36
arpu_3g_7                   73.36
total_rech_data_7           73.36
count_rech_3g_7             73.36
arpu_2g_7                   73.36
av_rech_amt_data_7          73.36
av_rech_amt_data_8          73.24
count_rech_2g_8             73.24
arpu_2g_8                   73.24
fb_user_8                   73.24
date_of_last_rech_data_8    73.24
arpu_3g_8                   73.24
total_rech_data_8           73.24
count_rech_3g_8             73.24
max_rech_data_8             73.24
night_pck_user_8            73.24
```

```
0    92.114094
1     7.885906
Name: churn, dtype: float64
```



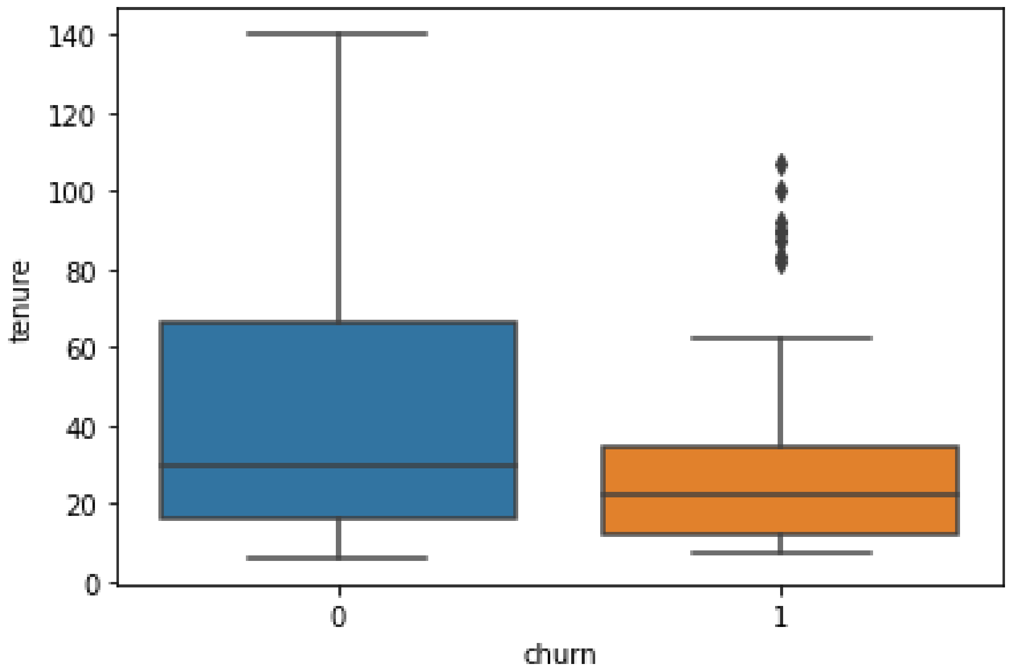
As we can see that 91% of the customers do not churn, there is a possibility of class imbalance

Since this variable churn is the target variable, all the columns relating to this variable(i.e. all columns with suffix _9) can be dropped forn the dataset.



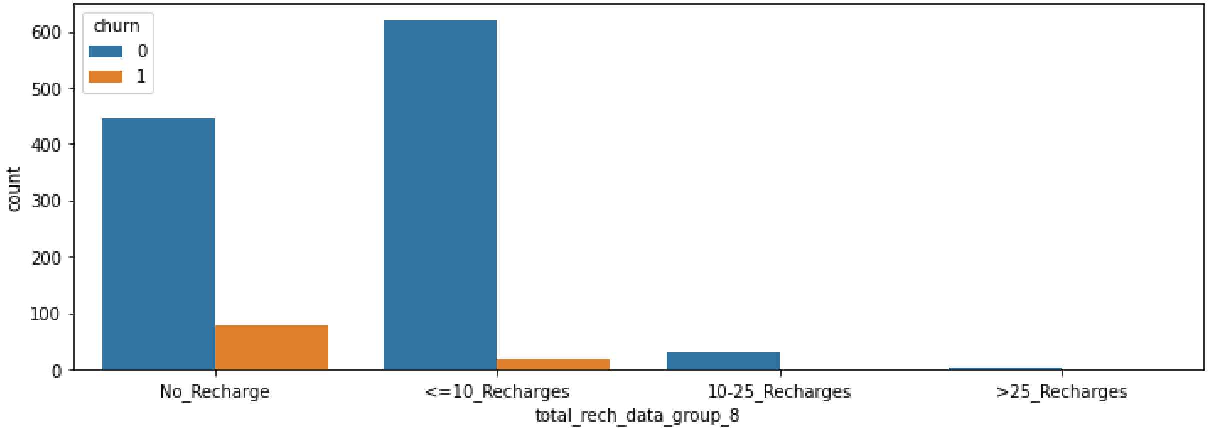
- It can be seen that the maximum churn rate happens within 0-6 month, but it gradually decreases as the customer retains in the network.

The average revenue per user is good phase of customer is given by arpu_6 and arpu_7. since we have two seperate averages, lets take an average to these two and drop the other columns.

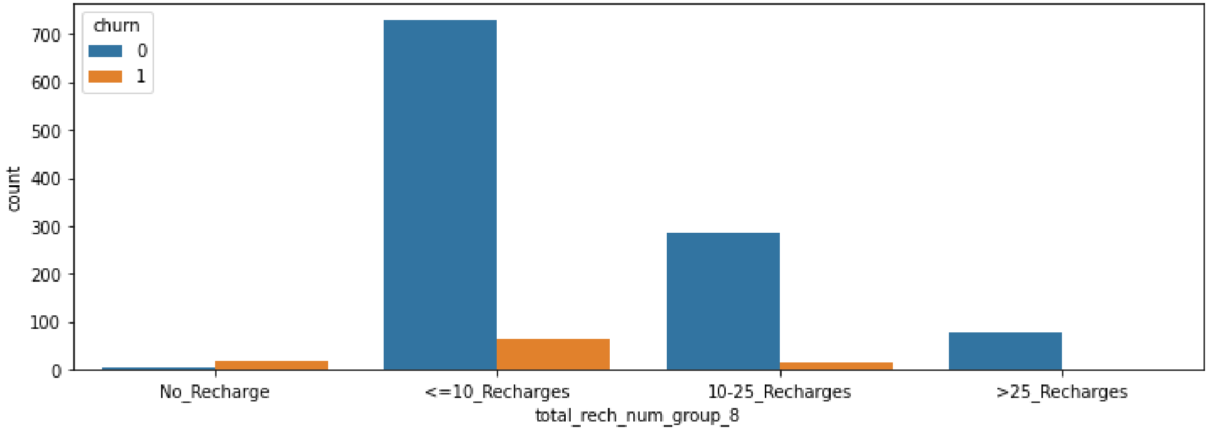


- From the above plot , its clear tenured customers do no churn and they keep availing telecom services

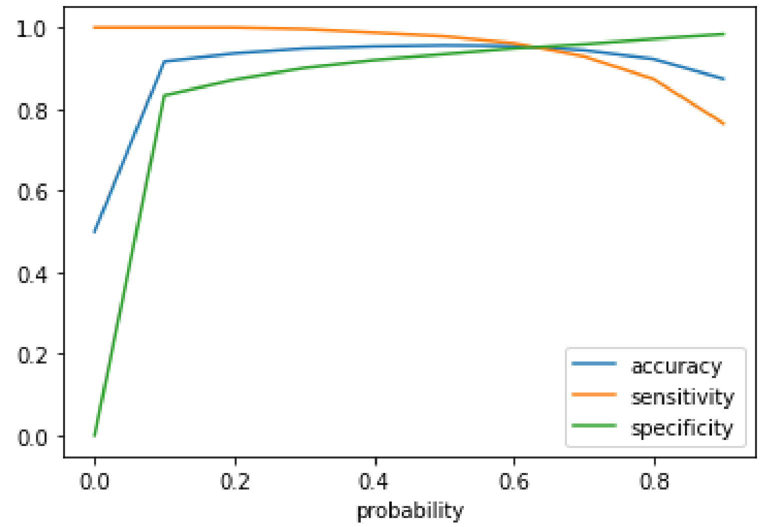
```
Name: total_rech_data_group_8, dtype: int64
```




```
                    Distribution of total_rech_num_8 variable
 <=10_Recharges     795
10-25_Recharges    298
>25_Recharges       79
No_Recharge         20
Name: total_rech_num_group_8, dtype: int64
```

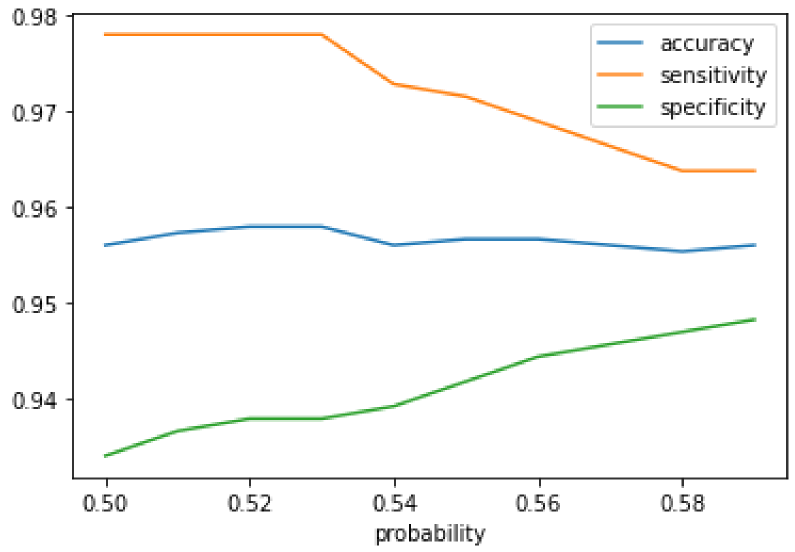


As the number of recharge rate increases, the churn rate decreases clearly.

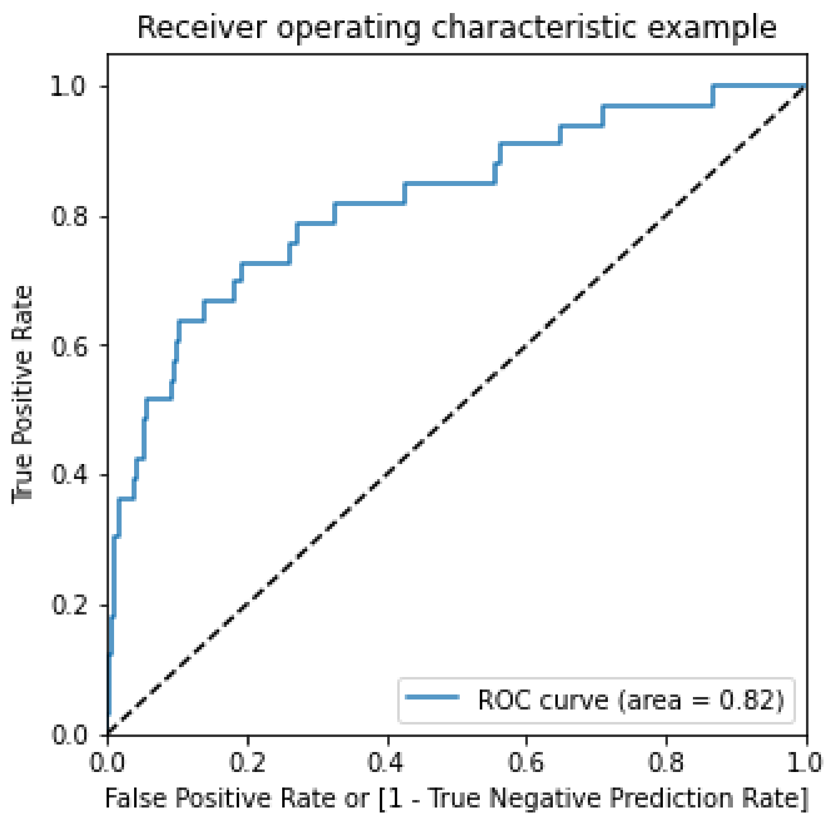


- Initially we selected the optimm point of classification as 0.5.

From the above graph, we can see the optimum cutoff is slightly higher than 0.5 but lies lower than 0.6. So lets tweek a little more within this range.



From the above graph we can conclude, the optimal cutoff point in the probability to define the predicted churn variabe converges at 0.54



The AUC score for train dataset is 0.90 and the test dataset is 0.87.

This model can be considered as a good model.

Model analysis

- 1.We can see that there are few features have positive coefficients and few have negative.
- 2.Many features have higher p-values and hence became insignificant in the model.
- Coarse tuning (Auto+Manual)
- We'll first eliminate a few features using Recursive Feature Elimination (RFE), and once we have reached a small set of variables to work with, we can then use manual feature elimination (i.e. manually eliminating features based on observing the p-values and VIFs).

Recommendations

- Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.
- Customers decreasing monthly 2g usage for August are most probable to churn.
- Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn.
- roam_og_mou_8 variables have positive coefficients (0.7135). That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.

- Target the customers, whose minutes of usage of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August).
- Target the customers, whose outgoing others charge in July and incoming others on August are less.
- Also, the customers having value based cost in the action phase increased are more likely to churn than the other customers. Hence, these customers may be a good target to provide offer.
- Customers, whose monthly 3G recharge in August is more, are likely to be churned.